PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68, C12N 15/10 — A1

(11) International Publication Number: WO 00/23622

(43) International Publication Date: 27 April 2000 (27.04.00)

(21) International Application Number: PCT/US99/23906

(22) International Filing Date: 15 October 1999 (15.10.99)

(30) Priority Data:
09/174,328 16 October 1998 (16.10.98) US

(71) Applicant: VALIGENE CORPORATION [US/US]; Suite 2300, 70 East 55th Street, New York, NY 10022 (US).

(72) Inventors: IRIS, Francois, J.-M.; 3, rue du Bouquet, F-92370 Chaville (FR). POURNY, Jean-Louis; 150, rue Perronet, F-92200 Neuilly (FR).

(74) Agents: ANTLER, Adriane, M. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS FOR MANIPULATING COMPLEX NUCLEIC ACID POPULATIONS USING PEPTIDE-LABELED OLIGONUCLEOTIDES

(57) Abstract

The present invention relates generally to methods of labeling, sorting and screening populations of nucleic acids. More particularly, the present invention relates to a method for sorting and comparing complex populations of nucleic acid, such as cDNA libraries. These complex populations of nucleic acid may be derived from cells or tissue types having variations in phenotype of potential clinical interest. The method is referred to generally as the ValiGene^{SM} Peptide- Labeled Oligonucleotide method, or VG-PLO^{SM}, and involves the use of distinguishable and identifiable peptide tags linked to identical oligonucleotide primers.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHODS FOR MANIPULATING COMPLEX NUCLEIC ACID POPULATIONS USING PEPTIDE-LABELED OLIGONUCLEOTIDES

1. FIELD OF THE INVENTION

The present invention relates generally to methods of labeling, sorting, comparing and isolating populations of nucleic acids. More particularly, the present invention relates to a method for sorting and comparing complex populations of nucleic acid, such as cDNA libraries. These complex populations of nucleic acid may be derived from cells or tissue types having variations in phenotype of potential clinical interest. The methods are referred to generally as ValiGene℠ Peptide-Labeled Oligonucleotide methods, or VG-PLO℠, and involve the use of distinguishable and identifiable peptide tags linked to oligonucleotide primers to manipulate nucleic acids.

2. BACKGROUND OF THE INVENTION

Labeled oligonucleotides have been used for detection of specific sequences. For example, Burdick and Oakes (Diagnostic Kit and Method Using a Solid Phase Capture Means For Detecting Nucleic Acids, European Patent Publication No. EP 0370 694 A2, Date of publication May 30, 1990) disclose the use of oligonucleotide primers, labeled with a label, with specific nucleic acid sequences which are complementary to a predetermined sequence of interest. This method is limited to the identification of a known sequence within a given sample and each pair of primers must correspond to a single predetermined PCR product.

Several gene expression assays are now becoming practicable for quantitating the effect of a drug on expression of a large fraction of the genes and proteins in a cell culture (*see, e.g.*, Schena et al., 1995, Quantitative Monitoring of Gene Expression Patterns with a Complimentary DNA Micro-array, Science 270:467-470; Lochort et al., 1996, Expression Monitoring by Hybridization to High-density Oligonucleotide Arrays, Nature Biotechnology 14:1675-1680; Blanchard et al., 1996, Sequence to array: Probing the genome's secrets, Nature Biotechnology 14, 1649; 1996, U.S. Patent 5,569,588, issued October 29, 1996 to Ashby et al. entitled "Methods for Drug Screening"). Raw data from these gene expression assays are often difficult to coherently interpret. Such measurement technologies typically return numerous genes with altered expression in response to a drug, typically 50-100, possibly up to 1,000 or as few as 10.

Few methods exist to rapidly compare multiple mixtures of nucleic acids to select genes that are differentially expressed or are shared by some but not all phenotypes-of-interest. Accordingly, there is a need in the art for methods which rapidly and efficiently allow comparison among nucleic acid populations.

3. SUMMARY OF THE INVENTION

This invention provides methods of labeling, sorting, comparing and screening multiple, complex populations of nucleic acids. These populations may be cDNA libraries constructed from phenotypically distinguishable cell or tissue types of interest. The method is extremely flexible and is adaptable to perform numerous complex sorting and comparing tasks.

The methods of this invention may also be used to increase and supplement the analytical powers of other techniques of manipulating complex cDNA population. A major advantage of the methods of this invention is the ability to screen multiple populations of cDNAs derived, for example, from different tissues belonging to the same individual or to phenotypically different cell types present concurrently within a given tissue sample.

This invention takes advantage of the ability to follow multiple populations of nucleic acids through various sorting and molecular comparison procedures. The invention employs oligonucleotide primers having distinguishable and identifiable peptide tags, which primers can be used to prime PCR reactions from vector sequences. Using such oligonucleotide primers, inserts from any given cDNA library can be labeled with library-specific peptide tags. The distinguishable tags serve to identify the library-of-origin of any given insert. Further, the distinguishable tags can be used to selectively sort and isolate inserts based on their library of origin regardless of the complexity of the mixture of products. For example, one can use a chromatography matrix having an antibody specific to one of the distinguishable peptide tags. Such a matrix can trap or retain fragments, both single and double-stranded, which bear the specific peptide tag. Fragments which do not contain this tag will be left free in the flow-through.

The methods of the invention make use of polymerase chain reaction (PCR) primarily to linearize all inserts within a given cDNA library and to affix a distinguishable

and identifiable peptide label to all inserts from a particular cDNA library so as to indicate their library-of-origin. PCR can also be used at various stages to amplify complex mixtures of products.

Nucleic acid sample populations may be derived from many different sources. Such sources may include different phenotypes present concurrently within a given tissue sample or different tissues belonging to the same individual. One phenotype may, but does not need to be, "healthy" and another typical of a disease state. The methods of the invention allow identification of genes that are specifically expressed in association with each phenotype as well as a comparison of genes which are expressed independently of the phenotype or are shared by some phenotypes but not all.

In a first embodiment this invention provides a method comprising the following steps: (a) labeling DNA from each of a plurality of cDNA libraries using PCR with oligonucleotide primers having a label unique to each library; (b) contacting DNA labeled in step (a) with a first said label with DNA labeled in step (a) with a different said label and (c) sorting DNA contacted in step (b) using one or more molecules, each molecule being capable of binding the label unique to each library.

This invention further provides in the first embodiment additional methods wherein the label unique to each library is a 5'-peptide label.

This invention further provides in the first embodiment an additional method wherein the label unique to each library is biotin.

This invention further provides in the first embodiment additional methods wherein the one or more molecules is an antibody.

This invention further provides in the first embodiment, methods wherein the oligonucleotide primers prime PCR from vector sequences common to the nucleic acids within a particular library (thus a different one such primer is used for each library) or the oligonucleotide primer primes PCR from vector sequences common to the plurality of cDNA libraries (thus the same oligonucleotide primer is used for priming PCR for the entire plurality).

This invention further provides in the first embodiment a method of sorting which comprises (d) denaturing hybrid DNA strands resulting from step (b); (e) contacting single strands denatured in (d) with single strand binding protein to prevent strand

reannealing; and (f) contacting the single strand binding protein coated single strands formed in (e) with one or more molecules, each molecules being capable of binding one of the labels unique to each library.

This invention further provides in the first embodiment a method wherein at least one of the one or more molecules in (e) is an antibody.

This invention provides in a second embodiment a method of cDNA library comparison comprising: (a) labeling DNA from a first cDNA population by PCR using oligonucleotide primers which have a first 5'-peptide label; (b) labeling DNA from a second cDNA population by PCR using oligonucleotide primers having a second 5'-peptide label; (c) contacting DNA labeled in step (a) with DNA labeled in step (b) under conditions such that hybridization can occur and (d) separating DNA having the first and the second 5' peptide labels from DNA having only the first or the second 5' peptide label.

This invention further provides in the second embodiment additional methods wherein the first cDNA population is from one or more cells or an organism subjected to a first condition and the second cDNA population is from one or more cells or an organism of the same type not subjected to said first condition.

This invention further provides in the second embodiment additional methods wherein the first cDNA population is from one or more cells or an organism subjected to a first condition and the second cDNA population is from one or more cells or an organism of the same type subjected to a second condition.

This invention further provides in the second embodiment additional methods wherein the first and second cDNA populations are from cells or organisms that differ phenotypically.

This invention further provides in the second embodiment additional methods wherein the nucleotide sequences of the oligonucleotide primer pair having the first 5'-peptide label and the nucleotide sequences of the oligonucleotide primer pair having the second 5'-peptide label are the same.

In a third embodiment this invention provides a method of monitoring gene expression comprising: (a) contacting mRNA from a cell with an RNA-dependent DNA polymerase and a 5'-dephosphorylated target-specific primer (i.e., specific to the gene of which it is desired to monitor expression); (b) contacting any DNA:RNA hybrids

synthesized in step (a) with a nuclease to remove single-stranded RNA extensions; (c) after step (b) ligating the DNA:RNA hybrids molecules to a partly double-stranded phosphorylated second primer (e.g., a primer that is not target specific); (d) labeling products ligated in step (c) by PCR with a first primer complementary to the target-specific primer used in step (a), said first primer being labeled with a first label and a second primer complementary to one strand of the double-stranded phosphorylated second primer in (e), said second primer being labeled with a second label that is distinguishable from the first label; (e) contacting the PCR products labeled in step (d) with one or more molecules immobilized on a solid support capable of binding the first label; (f) washing the solid support; and (g) contacting the support washed in step (f) with one or more molecules capable of binding the second label.

This invention further provides in the third embodiment an additional method wherein the nuclease is mung-bean nuclease.

This invention further provides in the third embodiment an additional method wherein the partly double-stranded phosphorylated second primer is an M13 forward sequencing primer.

This invention further provides in the third embodiment an additional method wherein the first label is a peptide label.

This invention further provides in the third embodiment additional methods wherein at least one of the one or more molecules in step (e) is an antibody.

This invention further provides in the third embodiment additional methods wherein at least one of the one or more molecules in step (g) is streptavidin-linked horseradish peroxidase.

A fourth embodiment provides a method of identification of cDNA inserts represented in a first cDNA library and not represented in a plurality of other cDNA libraries comprising: (a) labeling DNA inserts from each cDNA library by polymerase chain reaction using oligonucleotide primers having a label unique to each library; (b) hybridizing DNA labeled in step (a); (c) contacting DNA hybridized in step (b) with a plurality of immobilized antibodies capable of recognizing the label unique to each of the plurality of other cDNA libraries but not the label unique to the first cDNA library; and (d) recovering DNA which is not bound by the plurality of immobilized antibodies.

This invention further provides in the fourth embodiment additional methods wherein the DNA hybridized from each of the plurality of other cDNA libraries is in excess relative to the first cDNA library. Furthermore, additional methods are provided which employ from a 2-fold to a 100-fold excess, from a 2.5-fold to a 10-fold excess and wherein the excess is a 3-fold excess.

This invention further provides in the fourth embodiment additional methods wherein the label unique to each library is a peptide label. Furthermore, methods are provided wherein the peptide label is 3 to 12 amino acid residues. Furthermore, methods are provided wherein the label is a thermophilic protein label.

This invention further provides in the fourth embodiment additional methods wherein one of the plurality of antibodies in step (c) is immobilized on a separate affinity column. Furthermore, methods are provided wherein the separate affinity columns are physically linked in series in any order.

This invention further provides in the fourth embodiment additional methods wherein the column flow-through is applied to the separate, physically-linked affinity columns one or more times. Furthermore, a method is provided wherein the column flow-through is applied to the separate, physically-linked affinity columns three times.

This invention further provides in the fourth embodiment a method wherein the DNA retained by the antibody specific for the label unique to the first cDNA library is recovered and cloned.

In a fifth embodiment this invention provides a method of identification of cDNA inserts represented in a first cDNA library and in a second cDNA library, and not represented in a plurality of other cDNA libraries, comprising: (a) labeling DNA from each cDNA library by PCR using oligonucleotide primers having a label unique to each library; (b) hybridizing DNA labeled in step (a); (c) contacting DNA hybridized in step (b) with a plurality of immobilized antibodies capable of recognizing the label unique to each of the plurality of other cDNA libraries but not the label unique to the first cDNA library or the second cDNA library; and (d) recovering DNA which is not bound by the plurality of immobilized antibodies.

This invention further provides in the fifth embodiment methods wherein DNA hybridized from each of the plurality of other cDNA libraries is in excess relative to

the first and second cDNA libraries. Furthermore methods are provided wherein the excess is from 2-fold to a 100-fold excess, wherein the excess is from 2.5 fold to a 10-fold excess and wherein the excess is a 3-fold excess.

This invention further provides in the fifth embodiment methods wherein the label unique to each library is a peptide label. Furthermore methods are provided wherein the peptide label is from 3 to 12 amino acid residues. Furthermore, methods are provided wherein the label unique to each library is a thermophilic protein label.

This invention further provides in the fifth embodiment methods wherein each of the plurality of antibodies in step (c) is immobilized on a separate affinity column.

Furthermore a method is provided wherein the separate affinity columns are physically linked in series any in order.

This invention further provides in the fifth embodiment a method wherein the column flow-through is applied to the separate, physically-linked affinity columns one or more times. Further a method is provided wherein the column flow-through is applied to the separate, physically-linked affinity columns three times.

This invention further provides in the fifth embodiment methods wherein DNA recovered in step (d) is further contacted with an antibody specific for the label unique to the first cDNA library or the label unique to the second cDNA library so as to concentrate cDNA fragments specific to the first cDNA library and the second cDNA library. Furthermore, methods are provided wherein the concentrated cDNA fragments specific to the first cDNA library and the second cDNA library are recovered and cloned. In addition methods are provided wherein the concentrated cDNA fragments specific to the first cDNA library and the second cDNA library are separated. Furthermore, a method is provided whereby the separation is carried out by denaturation, coating with single-strand binding protein and contacting with an antibody specific for the label unique to the first cDNA library or the second cDNA library.

In a sixth embodiment this invention provides methods for matrix analysis of a plurality of cDNA libraries comprising: (a) labeling cDNA inserts from each of the plurality of libraries with a distinguishable label; (b) hybridizing cDNA inserts labeled in step (a); (c) contacting cDNA inserts hybridized in step (b) with an affinity column capable of binding a distinguishable label; and (d) eluting the affinity column.

This invention further provides in the sixth embodiment additional methods wherein the distinguishable label is a peptide label, and the step of labeling comprises priming PCR from cDNA library vector sequences by use of an oligonucleotide primer pair having said peptide label attached to the 5 ends of said primer pair.

This invention further provides in the sixth embodiment a method wherein the labeled cDNA fragments from each library are hybridized in equal proportions.

This invention further provides in the sixth embodiment methods wherein the affinity column capable of binding a distinguishable label is an antibody affinity column. Furthermore a method is provided wherein the antibody-affinity column is eluted with a pH gradient.

This invention further provides in the sixth embodiment a method wherein eluted DNA is denatured to separate strands originating from two different libraries. Furthermore, a method is provided wherein the denatured strands are isolated by: (a) coating with single-strand binding protein and (b) contacting with an affinity column capable of binding a distinguishable label.

The methods of this invention may also be used in another embodiment to construct subtracted cDNA libraries. The sequences obtained from any of the above-described procedures may be used to remove a homologue from libraries known to share such homologue or from any given unknown library. The library to be subtracted is present as purified double-stranded clones. This embodiment utilizes the ability of *E. coli* RecA protein to form stable triple-stranded structures between homologous sequences. Such triple-stranded structures are present as RecA coated single-stranded filaments and double-stranded linear and circular duplexes. The method of this embodiment comprises: (a) amplifying and labeling sequences identified as being shared by different libraries by PCR using vector-specific primers with 5'-peptide tag; (b) denaturing the tagged sequences; (c) cooling (*e.g.*, flash freezing) the denatured products to prevent renaturation; (d) adding an aliquot of RecA protein and non-hydrolyzable ATP (*e.g.*, ATPγS); (e) thawing the mixture; and (f) adding an aliquot of the library to be subtracted in the form of closed circular clones (*e.g.* plasmids or phagemids).

4. BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Schematic representation of the results obtained from Phase I, Phase II and Phase III, respectively, of the antibody affinity columns used for sorting labeled cDNA fragments.

FIG. 2. Schematic representation of the cDNA fragments comprising the input and output of Phase IV of the sorting process.

FIG. 3. RNA:DNA hybrid produced by cDNA first-strand synthesis, including the target-specific primer on the 5' end of the cDNA strand.

FIG. 4. Ligation of the partly double-stranded standard primer of the RNA:DNA hybrid, including partial ligation to RNA strand only.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods referred to generally as ValiGene℠ Peptide-Labeled Oligonucleotide methods (VG-PLO℠ methods) for manipulating (*e.g.*, labeling, sorting, isolating and/or screening) two or more complex populations of nucleic acids. Such nucleic acids may be derived from a variety of sources, typically cDNA libraries representing different phenotypes. For example, the cDNA libraries used may represent phenotypes present (i) concurrently within a given tissue (*e.g.* normal and cancerous portions of a biopsy specimen) (ii) within different tissues belonging to the same or different individuals, (iii) among different cell lines, or (iv) within the same cell line subjected to one or more different treatments.

In the methods of this invention generally, polymerase chain reaction (PCR) is employed to linearize inserts within a given cDNA library and to affix a distinguishable and identifiable label to all inserts from the given cDNA library so as to indicate the library (and thus cell type, tissue or organism) of origin. In a preferred embodiment, oligonucleotide primers to which peptide labels are linked are used to prime PCR reactions from vector sequences of cDNA libraries.

Throughout this application reference is made to peptide labels and antibodies for binding said labels. In addition to peptide-antibody combinations, it will be understood by those skilled in the art that any label can be used, in combination with a suitable binding partner. Examples of such labels and binding partners include, but are not limited to, digoxigenin-antidigoxigenin, biotin-streptavidin, ligand (e.g. hormone)-receptor and carbohydrate-lectin combinations.

As used herein, the complexity of nucleic acid populations or mixtures will be understood by one of ordinary skill in the art to refer generally to the number of distinguishable clones in any given cDNA library or mixture of libraries. The complexity of nucleic acids analyzed by the methods of the invention may vary over a very broad range. Generally, there is no upper or lower limit on the complexity of a population or mixture to be analyzed. For example, in one embodiment the complexity of a population or mixture may be from 10 to 10,000,000. Further, the complexity may be from 100 to 1,000,000. Still further, the complexity may be from 500 to 500,000. In a preferred embodiment, the complexity of a mixture analyzed is about (± 20%) 150,000. In another preferred embodiment, the mixture of complexity (± 20%) 150,000 comprises five libraries, each library having a complexity of about 30,000. In another specific embodiment, the complexity of the population being analyzed is at least 10^3, 10^4, 10^5, or 10^6.

The methodology of the invention utilizes library-specific labels linked to primer pairs capable of recognizing vector sequences of the vector used to construct the library. In this way, all nucleic acid fragments generated by PCR amplification with such primers have identical vector sequences at their 5' and 3' ends, yet also have a distinguishable label indicating the library-of-origin. Described below are methods for: sorting cDNA fragments to isolate those distinguishable to a single cDNA library; sorting cDNA fragments to isolate those common to two libraries; sorting cDNA fragments to isolate those common to multiple but not all libraries; and sorting cDNA fragments to isolate fragments shared only by two libraries out of all libraries analyzed. In addition, set forth below are methods to construct subtraction libraries to monitor gene expression events, and to isolate full-length transcripts of partial length sequences, such as expressed sequence tags.

5.1 IDENTIFICATION OF cDNA FRAGMENTS SPECIFIC TO ONE OF A PLURALITY OF cDNA LIBRARIES

In this embodiment, the inserts from each of a plurality of cDNA libraries (*e.g.* A, B, C, D and E) are each linearized and tagged by PCR with a distinguishable label (*e.g.*, a peptide tag that comprises an epitope) attached to vector-specific primers. The nucleotide sequences of a vector-specific primer pair can be identical among libraries. However, each label (tag) is specific to a library-of-origin. In this way, all fragments produced have (a) identical nucleotide sequences at their 5' and 3' ends corresponding to the vector-specific primer sequences and (b) a label indicating the library-of-origin. The labeled inserts are then purified, *e.g.*, by exclusion chromatography, to remove all reaction components, including excess peptide-labeled primers. These purified, tagged PCR products are then combined, heat-denatured and allowed to reanneal.

The conditions under which renaturation and hybridization are carried out can vary. In a preferred embodiment of this invention, the combined, heat-denatured PCR products are maintained together at 98°C for ten (10) minutes. The solution is then allowed to cool from 98°C to 85°C over a period of five (5) minutes. The temperature of the mixture is maintained at 85°C for ten (10) minutes and then cooled to 65°C over a period of fifteen (15) minutes. The solution is then maintained at a temperature of 65°C for a further time period of fifteen (15) minutes. At this point, the reannealing process is considered complete.

Here also, the quantity of each PCR reaction product used in the hybidization reaction can vary. Since isolation of cDNA fragments specific to one library is desired, the other libraries are used in excess (*i.e.*, as a "mop"). Specifically, where one wishes to isolate fragments present in library A but not in B, C, D and E, one uses excess B, C, D and E. The amount of excess B, C, D, and E used will determine the efficiency of removal. In a preferred embodiment, 3-fold excess of B, C, D and E is used. In another embodiment, from 2-fold to 100-fold excess B, C, D and E is used. In yet another embodiment, from 2.5 to 10-fold excess B, C, D and E is used. Since the function of using excess B, C, D and E is to act as a "mop" for removal of homologous cDNA fragments from the library-of-interest (in this example, library A), there is no restriction on the upper limit of the excess

that can be used. However, a 3-fold excess will efficiently remove the fragments from library A that will form hybrids with B, C, D and E without being wasteful.

Aliquots of the reannealed mixture are then contacted with one or more solid phases, *e.g.*, by passage through separate chromatography columns, having a binding partner, preferably antibodies (Abs), specific for tags B, C, D and E, respectively. Alternatively, a single solid phase, *e.g.*, column, having antibodies specific for all four tags may be used. Methods of making Ab affinity columns are well known. For example, agarose beads (*e.g.*, Sepharose™ or Sepharose CL, Pharmacia) may be activated by use of carbonyldiimidazole or cyanogen bromide for Ab attachment. The beads are washed with dioxane in water and incubated at room temperature with carbonyldiimidazole. After incubation the beads are again washed with dioxane. The purified solution of the desired antibody may then be added to the activated beads and mixed overnight at room temperature. The beads are then washed with 1 M NaCl and enthanolamine is added. The beads are then ready for binding to the antigen, or may be stored. Many other methods for preparing antibody affinity columns are known to those skilled in the art (see especially, Antibodies - A Laboratory Manual, Harlow, Ed. Lane, D., Cold Spring Harbor Laboratory Press, 1988, pp. 519-540).

Many different antibodies can be used in the single and multiple antibody-affinity columns used in the various embodiments of this invention. In a preferred embodiment, an antibody is used that specifically binds to a short peptide of from 6 to 12 amino acids that is used as the label/tag. However, peptides of any length can be used as "tags" if the chosen peptide is known to spontaneously renature following heat denaturation (e.g. thermophilic proteins). In a preferred embodiment, the antibody releases the retained peptide label when placed in a weakly acidic solution (*e.g.*, pH 5.5). Elution of antibody-affinity columns on the basis of pH gradient is another preferred embodiment of this invention.

The flow-through can be applied to each B, C, D and E column or other solid phase one or more times. Multiple times is preferred. After several cycles, all or most of the fragments bearing B, C, D or E tags, on either single or double-strands, will be trapped. In a most preferred embodiment, the flow-through is applied to the column three times. The

flow-through will generally contain single and double-stranded fragments bearing the A tag only.

The temperature for running the antibody affinity columns may vary. In a preferred embodiment the temperature is from 4°C to 50°C. In a most preferred embodiment, the antibody-affinity columns are water-jacketed and maintained at a temperature of 37°C.

The B, C, D and E columns are next washed with a low-salt (*e.g.*, 50mM NaCl) buffer (*e.g.*, phosphate buffer). The flow-through and washes are pooled. The pooled flow-through and washes from the B, C, D, and E columns are then passed through a column containing an A-specific antibody only. The trapped A-tagged fragments can then be eluted, precipitated and amplified using PCR with unlabeled, vector-specific primers. In a preferred embodiment, the A-tagged fragments are amplified using 20 cycles of PCR and cloned for analysis. These cloned fragments are highly enriched for fragments specific to library A (*i.e.*, fragments not found in library B, C, D, or E).

Any method known in the art may be used to elute peptide-labeled fragments from antibody affinity columns used in the various embodiments of this invention. In a preferred embodiment, as mentioned above, columns are eluted by changing pH (pH gradient). In a most preferred embodiment, the antibodies, or other solid phase, chosen will release the peptide label in a weakly acidic pH (*e.g.*, 5.5).

The same series of steps can be used to isolate cDNA fragments tagged with B, C, D or E peptides by changing the "mop". For example, to isolate the fragments labeled with B, one would start with an A, C, D and E multi-antibody affinity column to retain everything but B-labeled fragments. The pooled flow-through and washes of this column would then be passed over a column containing only a B-specific antibody. The same approach can be used to isolate fragments specific to the C, D and E libraries.

In the example above, where A-specific fragments are isolated, the material trapped in the first column (multi-antibody column) will contain all fragments tagged with B, C, D and E labels. This will include hybrids containing one A-tagged strand. This material can also be eluted and further sorted using other embodiments of the invention, or as otherwise desired by the practitioner.

5.2 IDENTIFICATION OF cDNA FRAGMENTS SPECIFIC TO TWO OR MORE OF A PLURALITY OF cDNA LIBRARIES

In this embodiment, the methods of this invention are used to isolate cDNA fragments that are common to, (*i.e.*, that will form hybrids with) cDNA fragments from one or more of the other libraries. For example, if one starts with a plurality of cDNA libraries (*e.g.*, A, B, C, D and E), this embodiment allows the isolation of transcripts common to the A and C libraries (or to any other two specified libraries). In this example, by way of explanation is described the isolation of cDNA fragments from the A library which form hybrid duplexes with cDNA fragments from the C library.

In this embodiment, as in Section 5.1 above, the inserts from each of a plurality of cDNA libraries (A, B, C, D and E) are linearized and tagged by PCR with a distinguishable library-specific label attached to vector-specific primers. Accordingly, as above, all fragments produced have identical nucleotide sequences at their 5' and 3' ends corresponding to the vector-specific primer sequences, and a distinguishable label indicating the library-of-origin. The labeled PCR products are then purified by exclusion chromatography to remove all reaction components, including excess labeled primers. The purified, labeled PCR products are then combined, heat-denatured and allowed to reanneal.

In this embodiment, as in the first approach, the quantity of each PCR reaction product used in the hybridization can vary. In this example, where one is interested in isolating A:C hybrids, one would use an excess of libraries B, D and E. The purpose of this excess is to act as a "mop" as in the previous approach. There is no restriction on the upper limit of the excess of B, D and E. However a 3-fold excess will efficiently remove cDNA fragments from both the A and C libraries that will form hybrids with any fragments from B, D and E libraries without being wasteful. Thus, a 3-fold excess is the preferred embodiment. In another embodiment, a 2-fold to 100 fold excess of B, D and E over A and C is used. In another embodiment, a 2.5 fold to 10 fold excess of B, D and E over A and C is used. The degree of excess of B, D and E will determine the efficiency of the "mop". An excess of less than 2-fold could be used, but significant quantities of A or C fragment which could hybridize to fragments in B, D or E may remain at the end of the procedure.

The reannealed mixture is then passed through a chromatography column containing antibodies specific to the labels on all the libraries except the two of interest. In

this example, the column would contain antibodies to labels on libraries B, D and E. The flow-through of this multi-antibody column can be applied one or more times. However, multiple times is preferred since each pass-through will increase the percentage of B, D or E-labeled fragments which will be retained in the column and therefore removed from the flow-through. After several cycles, all or most of the fragments which bear the B, D or E label will be removed and the flow-through will contain only fragments bearing the A or C label. These fragments will consist of A:A and C:C duplexes and, in addition, may contain A:C hybrids (*i.e.*, the fragments-of-interest). These A:C hybrids contain cDNA fragments from the A library which formed hybrid duplexes with fragments from the C library and were not removed by the "mop" of fragments from the B, D and E libraries.

The A:C hybrids are the product of interest. The mixture containing A:A and C:C duplexes and A:C hybrids is next passed through an antibody affinity column with immobilized anti-A label antibody. This column will retain all or most fragments which bear at least one A-label. This will include A:A duplexes and the A:C hybrids of interest. The flow-through may be applied to this single-antibody column one or more times. The amount of A-labeled fragments retained will be increased with each pass. In a preferred embodiment, the flow-through is passed through the column three times.

The material retained by this column is now eluted and the trapped material recovered and precipitated. The recovered material consists of A:A duplexes and A:C hybrids. These double-stranded fragments are then heat denatured.

Immediately after the denaturation, the resulting single strands are cooled rapidly to prevent renaturation. For example, this can be accomplished by rapidly cooling the heat-denatured material on a bath of dry-ice and methanol. Single-strand binding protein (SSB) is added to the frozen mixture. This protein will stabilize single DNA strands by coating them, thereby preventing renaturation. In this example, the SSB is then added in excess to the frozen mixture of denatured single-strands. This frozen mixture is then warmed to allow the SSB to enter the solution and contact the single strands. In a preferred embodiment, the mixture is heated from the temperature of the dry-ice/methanol bath to 37°C and maintained at that temperature for a few minutes (*e.g.*, between 5 and 10 minutes). The SSB will coat and stabilize the single DNA strands and will prevent reformation of hybrids and the formation of secondary structures.

The single strands of DNA consist of fragments from the A library that formed hybrids with other fragments from the A library, and fragments from the C library that formed hybrids with fragments from the A library. The C library fragments present at this stage will be limited to those fragments that were able to hybridize with A library fragments and were thus retained on the anti-A antibody column. In addition, these C library fragments did not hybridize with and thus were not removed by the excess of B, D and E fragments used as the "mop". Accordingly, these C library fragments are represented in the A library but not in the B, D, or E library.

In this embodiment, a further step is now employed to separate the SSB-coated A and C single strands. This step consists of passing the SSB-coated single-strands through an antibody-affinity column. This column may contain either immobilized anti-A antibody or immobilized anti-C antibody. If the anti-A column is used, then A-labeled fragments will be retained by the column and the C-tagged fragment will remain in the flow-through and washes. If the anti-C antibody column is used, then the C-labeled fragments will be retained by the column and may be eluted from this column. In either case, the A-labeled fragments and the C-labeled fragments can be recovered. The recovered fragments are then extracted to remove SSB, and PCR amplified. These C-labeled fragments may be cloned for further analysis or they can be used as pooled probes (*i.e.*, "subtraction probes") to remove their homologues from the original A or C libraries (*see e.g.*, Section 5.4 below).

In a variation of this embodiment of the invention, more than two libraries-of-interest may be designated. For example, if one is interested in fragments that may form hybrids between library A, B and C but not with libraries D and E, then the first step would employ an excess of D and E PCR reaction products over those of A, B and C. In this embodiment the multiple antibody column would contain anti-D and anti-E antibody. The excess of D and E used would form the "mop" to remove cDNA fragments that formed hybrids with any fragments in the A, B, or C libraries. The flow-through and washes of this multi-antibody column would contain A:A, B:B and C:C duplexes but would also contain A:B, A:C and B:C hybrids if any had formed. These hybrids are the products-of-interest in this embodiment. If this material is now passed through an anti-A column, then A:B and A:C hybrids will be retained. An anti-B antibody column will retain any A:B and B:C

hybrids, and an antibody column containing anti-C antibody will retain C:B and C:A hybrids. The material retained on these three columns may be eluted and isolated by the methods used above. This would consist of denaturating the double-stranded hybrids, rapid cooling on a dry-ice/ methanol bath followed by warming in contact with an excess of SSB. The SSB coated single-strands could now be isolated by passing through an antibody column containing a single antibody specific to either A, B, or C label. The fragments eluted from these columns would contain the cDNA fragments of interest.

5.3 MATRIX ANALYSIS OF A PLURALITY OF LIBRARIES

In another embodiment of the invention, an array or matrix comparison is made among a plurality of cDNA libraries constructed and manipulated, in part, using variations of the procedures set forth above. In this embodiment, any cDNA fragments present in two of N libraries can be isolated, where N is the total number of libraries subjected to the matrix analysis. Further, any cDNA fragments present in X of N libraries can be isolated, where X is the number of libraries in which the particular cDNA fragments isolated are found. Still further, any cDNA fragment present in *only* two (or three, or X) of N libraries can be isolated. Indeed, the cDNA fragments common to any desired number (X) of any number (N) of libraries analyzed can be isolated, and whether or not these fragments are exclusively shared among a subset of N libraries analyzed can be determined.

A major advantage of this embodiment is the absence of a necessity for having any knowledge of which genes (i.e. cDNA fragments) may or may not be represented in a given cDNA library before beginning the comparative analysis. Further, the degree or extent of homology (i.e. similarity) among cDNA fragments obtained from a plurality of libraries also need not be known. Still further, one need not know whether a specific library-of-interest shares any similar cDNA inserts with any other libraries prior to beginning the analysis. In summary, the results of a matrix analysis reveals which cDNA fragments are common (i.e. similar enough to hybridize) in any two or more of a plurality of libraries-of-interest.

To illustrate this embodiment, we again start with five cDNA libraries (A, B, C, D and E). First, the cDNA inserts of each library are separately linearized and labeled by PCR with a label distinguishable to each library. Again, a 5'-peptide label is preferred. As

above, the peptide label distinguishable to each library is attached to the 5' ends of an oligonucleotide primer pair used to prime PCR from library vector sequences.

After the PCR linearizing and labeling procedure, the contents of each labeled library is separately purified, *e.g.*, by exclusion chromatography. This step purifies the linearized, labeled inserts away from unwanted reaction components, such as excess peptide-labeled primers. This purification can be performed by any of the standard methods well known in the art (*e.g.*, PCR purification kit from Qiagen, Santa Clarita, California).

The purified and distinguishably labeled cDNA fragments from each of the five libraries are then mixed together, heat denatured and allowed to re-anneal. The relative proportions of material from each library mixed together in this reaction is determined by user discretion. The reaction may or may not employ an excess of material from one or more libraries over another one or more libraries. In a preferred embodiment, the labeled cDNA fragments from each library are mixed together in equal proportions.

As in the methods of the invention already described, this embodiment performs a comparative analysis of cDNA libraries based on hybridization and sorting of labeled cDNA inserts. Here, however, the analysis employs up to four stages or Phases of affinity columns, or any solid phase, capable of binding specific library labels. As before, any label known in the art suitable for labeling cDNA strands by PCR may be used. Further, any affinity column, or any solid phase, known in the art, capable of binding such labels may be used. In a preferred embodiment, the affinity columns are antibody-affinity columns capable of binding peptide labels.

The precise number of Phases employed in this embodiment, and their order of use, is determined in part by the result desired by the user. In this regard, all products of the comparative array created need not be analyzed and may be stored. For example, where one wishes to isolate inserts shared between any two libraries-of-interest and one is not concerned with isolating inserts exclusively present in these two libraries, then the analysis need only proceed through Phases I and II. However, where one wishes to isolate fragments exclusively present between libraries within a library Group, then Phases I, II and III are employed. Further, where one wishes to isolate fragments exclusively present in libraries from different library Groups, then Phases I, II and IV are employed. A library "Group" is defined as the eluent obtained from a Phase I column, as further set forth in the sections

below. The following narrative describes the steps employed to achieve the results just described. As noted above, while described in terms of antibody affinity columns and peptide labels, other types of solid phases with other types of binding partners to other types of label, can be used.

5.3.1 PHASE I

In Phase I, the re-annealed mixture of labeled cDNA fragments is applied sequentially to a series of antibody affinity columns, each column having an immobilized antibody capable of recognizing only one of the library labels. Alternatively, the re-annealed mixture may be divided into aliquots and applied to each column individually. In this five-library example, five columns are used. The order of application of the re-annealed mixture to the five individual columns can be any order. For example, the order can be A column, B column, C column, D column and E column, respectively. In a preferred embodiment, the five columns are physically linked in series. This arrangement has the advantages of efficiency in running the columns, of minimizing the volume applied to the columns, and of reducing losses of column flow-through and washes. Where the series of Phase I columns is physically linked, the order of the columns in the series is again not important and can be any order.

Each column in the Phase I series will trap or retain cDNA fragments having one specific label. Thus, the A column will retain all A-labeled fragments, the B column will retain all B-labeled fragments, etc. Fragments retained, for example, by the A column consist of A-labeled hybrids (i.e. A:A, A:B, A:C, A:D and A:E duplexes) and single-stranded A-labeled DNA.

Each of the five Phase I columns is next eluted individually. Where columns A through E were physically linked for application of the cDNA mixture and washes, they are dis-assembled prior to elution. The material obtained from elution of each Phase I column defines a separate library Group, as follows:

Library Group A	
from A column -	double-stranded A:A, A:B, A:C, A:D and A:E duplexes, and single-stranded A-labeled DNA;

Library Group B	
from B column -	double-stranded B:A, B:B, B:C, B:D and B:E duplexes, and single-stranded B-labeled DNA;
Library Group C	
from C column -	double-stranded C:A, C:B, C:C, C:D and C:E duplexes, and single-stranded C-labeled DNA;
Library Group D	
from D column -	double-stranded D:A, D:B, D:C, D:D and D:E duplexes, and single-stranded D-labeled DNA; and
Library Group E	
from E column -	double-stranded E:A, E:B, E:C, E:D and E:E duplexes, and single-stranded E-labeled DNA.

In this example, the material-of-interest is duplex DNA formed from hybridization of strands originating in two different libraries. Where five libraries are used for the input mixture of Phase I, as in this example, twenty different duplexes-of-interest may be formed. For example, in column A, trapped library Group A duplexes-of-interest consist of A:B, A:C, A:D and A:E duplexes. In column B, trapped library Group B duplexes-of-interest consist of B:A, B:C, B:D and B:E duplexes, etc. See FIG. 1, Phase I, for a complete listing of the array of products produced. Duplexes having identical labels on each strand and any single-stranded DNA trapped in Phase I columns is not generally of interest and is therefore not shown in FIG. 1.

5.3.2 PHASE II

The cDNA fragments (i.e. transcripts) shared between any two libraries-of-interest is isolated in Phase II. Here, the eluent from each of the five Phase I columns is rendered single-stranded prior to input over Phase II columns. This may be performed by any method known in the art. In a preferred embodiment, single-strand binding protein (SSB) is used. Thus, the eluent from each of the five Phase I columns (Groups A through E in FIG. 1) is separately heat-denatured and rapidly cooled (e.g. dry-ice/methanol bath). An excess of SSB is added, and each SSB-plus-DNA mixture is then warmed. In a preferred

embodiment, the temperture is increased to and maintained at 37°C for 10 to 15 minutes. SSB coats denatured single-strands and prevents renaturation and formation of secondary structures.

The output of each of the five Phase I single-antibody columns, now denatured and stabilized with SSB, is next applied to a series of Phase II columns. In a preferred embodiment, each series of Phase II columns is physically linked. Each Phase II column contains a single immobilized antibody specific for one of the distinuishable peptide labels used to label the input cDNA libraries. A Phase II series of columns may contain as many columns as the number of cDNA libraries N being analyzed. In an alternative embodiment, a Phase II series of columns contains N-1 columns, where the omitted column corresponds to the library Group (e.g. for library Group A, the A column may be omitted). The immobilized antibody in each Phase II column captures single strands bearing one of the distinguishable peptide labels. Each Phase II column is then separately eluted. In this way, an array of twenty groups of single-stranded cDNA fragments is isolated wherein each of the twenty groups contains fragments shared (i.e. hybridizable) between two libraries (see FIG. 1, Phase II).

For example, after elution of each Phase II column in the A library Group (N-1 approach), the following cDNA fragments are isolated in single-stranded form:

B column - cDNA fragments originating from the B library also present in the A library;

C column - cDNA fragments originating from the C library also present in the A library;

D column - cDNA fragments originating from the D library also present in the A library; and

E column - cDNA fragments originating from the E library also present in the A library.

As a further example, after elution of each Phase II column in the B library Group under the N column approach, the following cDNA fragments are isolated in single-stranded form:

B column - B cDNA fragments only;

A column - cDNA fragments originating from the A library also present in the B library;

C column - cDNA fragments originating from the C library also present in the B library;

D column - cDNA fragments originating from the D library also present in the B library; and

E column - cDNA fragments originating from the E library also present in the B library.

A similar list of isolated cDNA fragments can be constructed for each series of Phase II columns, thereby completing the array (see FIG. 1, Phase II).

Of course, any fragments in the array created by the output of the Phase II columns may be cloned for further analysis as desired by the user. Such fragments may also be used as "combination probes" for retrieval of corresponding double-stranded clones from an existing library using, for example, the RecA method detailed elsewhere herein. These fragments are also used as input to Phase III and Phase IV for isolation of cDNAs exclusively present in two or more desired libraries, either within or across library Groups, respectively, as further set forth below.

5.3.3 PHASE III

Phase III is used to isolate fragments shared exclusively between two or more designated members within a library Group. For example, Phase III allows isolation of fragments shared exclusively between libraries A and B, A and C, A and D, and A and E, within library Group A. Further, Phase III allows isolation of fragments shared exclusively between libraries B and A, B and C, B and D, and B and E, within library Group B. This pattern is equally applicable to library Groups C, D and E.

Phase III begins with single-stranded fragments eluted separately from each of a series of Phase II columns in a chosen library Group as described above. These fragments are first independently amplified by PCR and labeled with the relevant peptide label. For example, where library Group A is being subjected to a Phase III analysis, fragments eluted from the B column of Phase II are amplified using the B-specific peptide

label, fragments eluted from the C column of Phase II are amplified using the C-specific peptide label, etc.

All the independent members belonging to a given library Group are likewise amplified and labeled by PCR. The amplified products are then mixed, denatured and allowed to re-anneal. In this example (library Group A where libraries A through E are being analyzed), the reannealed Phase III input mixture is then divided into four aliquots. The number of aliquots needed depends on the number of separate labels in the mixture. In this example, the A library Group is analyzed and the labels used during the PCR amplifying process are B, C, D and E. The Phase III columns consist of four series of affinity columns, each series consisting of three single-antibody columns. Each of these four series of columns contain antibodies specific to three of the four labels used in the Phase III PCR amplification step. Where, as here, the A library Group is subjected to Phase III analysis, the four series of affinity columns contain:

Series 1 -	C, D and E antibodies;
Series 2 -	B, D and E antibodies;
Series 3 -	B, C and E antibodies; and
Series 4 -	B, C and D antibodies.

Phase III Series 1 retains any cDNA fragments labeled with C, D and E, allowing B-labeled duplexes and single strands to remain in the flow through. Within library Group A, these cDNA fragments are present in libraries A and B, but not in C, D or E. Therefore, cDNA fragments exclusively present in libraries A and B have been isolated. In the same manner, Phase III Series 2 allows only C-labeled fragments to pass, Phase III Series 3 allows only D-labeled fragments to pass, and Phase III Series 4 allows only E-labeled fragments to pass. Here, cDNA fragments exclusively present in libraries A and C, A and D, and A and E, respectively, have been isolated.

Thus generally, in each series column, one uses one column less than the number of labels used in the amplifying step. Further, one uses enough series to cover all different combinations of columns.

In an alternative embodiment, the flow-through of each of the four Series of multi-antibody columns just described above is next passed through another antibody column. These columns each contain a single antibody which is specific for the labeled

fragments allowed to pass in the Phase III Series columns. This step serves to concentrate the fragments, which otherwise might be difficult to recover from a large volume of flow-through and washes. The fragments retained by these four single-antibody columns are eluted and recovered. This material consists of concentrated cDNA fragments that are uniquely shared between two specific libraries. In this example, the fragments recovered are uniquely shared between libraries A and B, A and C, A and D, and A and E within library Group A.

5.3.4 PHASE IV

The output from Phase II can be further analyzed in Phase IV to determine whether cDNA fragments shared between any two libraries-of-interest in the array are distinguishable *across* library Groups rather than *within* library Groups. The Phase IV analysis thus complements the Phase III analysis by allowing one to ask essentially the same question using different input cDNA fragments. The user thus benefits by comparing the results of a Phase III analysis with the results of a Phase IV analysis. As in Phase III, the input DNA for Phase IV analysis is obtained from the output of Phase II. However, the labels attached in the PCR reactions prior to Phase IV analysis correspond to the library Group label and not to the original label of the fragment (see Box in FIG. 2).

Thus, importantly, the labels attached in the PCR prior to Phase IV analysis of Phase II products do *not* correspond to the library-of-origin of a particular fragment. Instead, the labels correspond to a library in which a given fragment has found a homolog (i.e. the library Group). In this way, an analysis similar to the Phase III analysis can be performed *across* library Groups.

For example, to perform a Phase IV analysis across library Groups, all fragments originating from A library and recovered from a B group column in Phase II are labeled with B peptide label (see Box in FIG. 2 at "Tag-B"). In a similar fashion, all fragments originating from A library which were recovered from a C, D, or E group column in Phase II are labeled with C, D and E peptide label, respectively (see Box in FIG. 2 at "Tag-C", "Tag-D" and "Tag-E", respectively).

After PCR amplification and labeling, all such differentially-labeled A-library fragments are mixed, denatured and allowed to re-anneal. The re-annealed mixture

is then divided into four aliquots, and each aliquot is passed over a multi-antibody affinity column similar to the Series 1-4 columns of Phase III. Thus, each multi-antibody column contains three label-specific antibodies. Where, as here, Phase IV analysis is performed for fragments originating in A library, four Series of Phase IV columns contain:

Series 1 - C, D and E antibodies;

Series 2 - B, D and E antibodies;

Series 3 - B, C and D antibodies; and

Series 4 - B, C and E antibodies.

As for Phase III analysis, the flow-through and washes from each Phase IV Series column may then be pooled and applied to a column containing a single antibody specific for the one label that remained untrapped. For example, the output of the Phase IV Series 1 column above would be pooled and passed over a column containing anti-B. Representative results for fragments originating in A library and for fragments originating in B library are shown in FIG. 4 (see "output of Phase IV"). As was the case described for the output of Phase III, the material eluted from each single-antibody column in Phase IV consists of concentrated cDNA fragments shared exclusively by two libraries (*i.e.*, not found in the other libraries of the analysis).

5.3.5 FURTHER CONSIDERATIONS

One can also isolate fragments common to three (or more) libraries, to the exclusion of others, by manipulating the Phase III and Phase IV Series columns to remove fewer fragments. For example, in a Phase IV analysis of fragments originating in library A directed to isolation of fragments present in A, C and E, but not in B or D, one could run a Phase IV Series column containing just B and D antibodies.

5.4 USE OF METHODS TO CONSTRUCT SUBTRACTED LIBRARIES

In another embodiment, the methods of this invention can be used to construct subtracted cDNA libraries, *i.e.*, to remove similar clones from two or more cDNA libraries. The method described herein takes advantage of the *E. coli* RecA protein's ability to form stable triple-stranded structures as recombination intermediates. RecA catalyzes a

homologous pairing and strand exchange reaction during *E. coli* homologous recombination. During the first step of this reaction, RecA coats a single strand of DNA and initiates an exchange reaction between the single strand and a homologous region of double-stranded DNA. A three-stranded nucleoprotein intermediate is formed, which, in the absence of ATP, is surprisingly stable (see West, 1992, Annu. Rev. Biochem., 61:603-640).

cDNA fragments-of-interest, such as those shared by different libraries identified as described above, are amplified by PCR using peptide-tagged vector-specific primers. Thus a distinguishable peptide tag marks a given set of cDNA sequences. Such sets of fragments are used concurrently as "subtraction probes". A subtraction probe set is purified by exclusion chromatography following PCR and heat-denatured. The cDNA mixture is then flash frozen (*e.g.*, dry-ice/methanol). An aliquot of RecA protein is added to the ice pellet along with non-hydrolyzable ATP (*e.g.*, ATP γS). The ice pellet is slowly thawed. Low temperature and the ATP analog prevent the RecA-bound single-stranded DNA from renaturing so that the subtraction probe remains single stranded and becomes coated with RecA.

The library to be subtracted, in the form of purified double-stranded, circular DNA, is added to the thawed pellet such that the RecA-coated single strands are present in large excess (20-50 fold). The mixture is heated to 37°C. The RecA-coated single strands scan the double-stranded cDNA library in search of homologous sequences, and pair with such sequences. Triple-stranded recombination intermediates are formed, although strand exchange will not occur due to the absence of a hydrolyzable form of ATP. The triple-stranded structures formed from single-stranded DNA and homologous double-stranded DNA are labeled with the specific peptide tag bound to the single strand. Such triple-stranded, labeled structures can now be separated from unlabeled, double-stranded circular molecules by passing the solution through a peptide tag-specific antibody column. Most or all clones corresponding to the labeled fragments will be removed from the library if the RecA-coated single strands are present in large excess over the plasmid clones.

The method of this embodiment is independent of the original nature of the nucleic acid used to construct the library. It can therefore be used with DNA libraries made from cDNAs or genomic DNAs.

In a preferred embodiment, both single-stranded fragments and double-stranded library plasmids share identical extremities (*i.e.*, 5' and 3' ends) over at least 10-15 bases, and the homologous fragments are at least 350 bp in length. If strong overall homology is present, perfect identity between fragments is not required for RecA to form stable triple-stranded structures (*see, e.g.*, Rao et al.,1995, Trends In Biological Science 20:109-113). In another preferred embodiment, the cloned inserts do not exceed 1-2 kilobase (kb) in length so that clones sharing only strong localized homologies with the subtraction probes are not selected.

5.5 USE OF METHODS IN THE MONITORING OF GENE EXPRESSION

In another embodiment of this invention, methods are provided to monitor gene expression events. Oligonucleotides labeled with specific and identifiable peptide labels are used, but in this embodiment the targets (*i.e.* genes) to be monitored for expression are known. These targets may belong to an expression cascade, for example, if the objective is to define the mechanism of action or physiological effects of a particular drug treatment. An alternative use for the methods of this embodiment is to monitor gene expression to define a phenotype based on the activation or repression of a specific phenotype-associated metabolic pathway. The advantage of this method is to provide a simple and rapid means to sort, separate and quantify the product (representing targets to be monitored for expression) based on the peptide label.

In addition, the methods of this embodiment allow a direct quantitative determination of the amount of target mRNA present. Briefly, a PCR reaction is carried out using unamplified cDNA from a first-strand synthesis reaction. For a fixed and limited number of PCR cycles (*e.g.*, from about 5 to 20 cycles), the product of the reaction is directly proportional to the initial amount of non-genomic, small-size (under 2kb) DNA target present in the reaction. The techniques of quantitative PCR are well known to those skilled in the art.

The methods of this embodiment will allow the direct monitoring of gene expression events, as well as the isolation of partial length transcripts, without the prior construction of the relevant cDNA libraries and starting from very small biopsy samples which would be too small to allow construction of new cDNA libraries.

The sensitivity and versatility of this embodiment allows its use to analyze the response of a specific phenotype to a given stimulus or set of conditions. In addition, this method provides a rapid and accurate means of directly determining the physiological effects of any form of treatment which affects gene expression, such as treatment with steroid hormones. Since this is done directly by looking at mRNA production, it is not necessary to wait for the overt clinical effects to show themselves or the production of serological factors. The method of this embodiment could therefore be used to make a rapid assessment of the probable effect of treatment, or to provide rapid and direct feed-back to allow therapeutic readjustments to be made to optimize outcome.

In this embodiment, total RNA is extracted from the tissue sample using standard methodologies well known to those skilled in the art. Total RNA is used for hybridization to target-specific probes. Each of these probes consists of a synthetic oligonucleotide labeled with a specific peptide epitope or tag at the 5' end and a fluorophore at the 3' end. The single-stranded probes are mixed with the samples of total RNA under conditions allowing hybridization of the probes to their target mRNA molecules, if present. Following hybridization, the mix is treated with a single-strand specific DNase in order to destroy all non-hybridized excess probes or to effect a separation between the peptide tag and the fluorescence label on probes remaining single-stranded. Here, one skilled in the art will recognize that other detectable labels may substitute for the fluorescence label. The mixture is then exposed to a solid surface onto which the tag-specific antibodies or other binding partners have been arrayed (e.g. an ELISA plate) hence identifying the relative position of each target-specific probe. Only those probes that have hybridized to their target will give rise to a fluorescence or other detecable signal at a specific location on the solid surface, the position in the array indicating the presence and identity of the target and the signal intensity indicating the relative abundance of each target within the original RNA sample.

This embodiment can also be used to isolate full- length forms of only partial-length transcripts. Total RNA is extracted as previously and aliquots of the total RNA are used for cDNA first-strand synthesis using target-specific, non-phosphorylated primers (*see*, FIG. 3).

The synthesis makes use of an RNA-dependent DNA polymerase (*i.e.*, reverse transcriptase) which does not possess RNase-H activity (such as Moloney murine leukemia virus reverse transcriptase). Methods for doing this are well known to those skilled in the art (Sambrook et al., 1989, Molecular Cloning A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press). The result of this synthesis is DNA:RNA hybrids with a target-specific primer on the 5' end of the DNA strand. Any RNA extensions can now be removed to produce blunt ends by treatment with mung-bean nuclease, which cleaves single-strand mRNA extensions.

This reaction can be performed under standard conditions for the use of mung bean nuclease. For example, the DNA:RNA hybrids may be suspended in a mung bean nuclease Buffer consisting of 50 mM sodium acetate (pH 5.0 at 25°C), 30 mM NaCl, 1 mM $ZnSO_4$. Mung bean nuclease in the amount of 1.0 unit per microgram of DNA:RNA hybrid is added and the mixture is incubated at 30°C for thirty (30) minutes. The enzymes may then be inactivated by phenol/chloroform extraction or by addition of SDS to 0.01%. The blunt-ended hybrids may be recovered by alcohol precipitation. Kowalski, D. et al. (1976) Biochemistry 15, 4457-4463; McCutchan, T.F. et al. (1984) Science 225, 626-628.

The sample is now purified by standard exclusion chromatography. After purification, the sample consists of the DNA:RNA hybrids together with the remainder of the total RNA species initially present. The exclusion chromatography removes the small RNA species (such as tRNA) and excess target-specific primer.

At this point, a ligation reaction is carried out using DNA ligase from T_4 bacteriophage. The ligase will catalyze the formation of phosphodiester bonds between adjacent 3'-hydroxyl and 5'-phosphate termini of DNA or RNA and will thus join the 3' end of a double-stranded DNA fragment to the 5' terminus of a double-stranded DNA:RNA hybrid molecule. The primer used is a partly double-stranded phosphorylated second primer (i.e. a primer that is not target-specific), for example, a M13 "forward" sequencing primer (see FIG. 4).

Bacteriophage T_4 DNA ligase will fully ligate the primer only to the phosphorylated end of the DNA:RNA hybrid. However, some of the primer molecules will also ligate to the 3' terminus of the RNA strand of the DNA:RNA hybrid. This will not affect the result because DNA polymerase enzyme in subsequent steps will not use RNA as

a template and because no template is available in the 3' direction and so priming at this site will not result in elongation by *de novo* synthesis.

T_4 DNA ligase purified from *E. coli* may be obtained from New England Biolabs (Waverly, Massachusetts). The reaction may be carried out in T_4 DNA Ligase Buffer which contains 50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, 25 microgram per milliliter bovine serum albumin. In a preferred embodiment, the rection is carried out at 16°C for between four (4) and sixteen (16) hours. Engler, M.J. and Richardson, C.C. (1982) in The Enzymes (Boyer P.D., ed.) Vol. 5, p. 3, Academic Press, San Diego, CA.

Following the ligation reaction, the sample is again purified by exclusion chromatography and amplified by PCR.

The PCR reaction makes use of both a peptide-tagged primer complementary to the target-specific primer previously used, and a biotinylated primer complementary to the partly double-stranded phosphorylated standard primer used in the ligation reaction. In a preferred embodiment, this PCR reaction includes 50 nanograms of yeast RNA per 30 microliters of solution. The number of cycles in this PCR reaction can vary. In a preferred embodiment, 20 or fewer cycles is used.

In a variation of this embodiment, the sample can be treated with RNase immediately following the ligation reaction and prior to PCR. This will destroy all RNA strands, including single strands of total RNA and the RNA strands of the DNA:RNA hybrid molecule. The sample can then be purified by exclusion chromatography, and PCR amplified and labeled as above. The amplified product is then purified by exclusion chromatography to remove all excess primers.

The amount of product produced by the PCR reaction can now be quantified by a modification of an enzyme-linked immunoassay technique (ELISA). The purified reaction mixture may be analyzed in microtiter wells coated with an antibody specific to the peptide label that was attached to the target-specific primer. Streptavidin-linked horseradish peroxidase can then be added to bind to the biotin moiety attached to the standard primer of the retained PCR products. A horseradish peroxidase substrate can then be added, and the reaction product quantified (see e.g. Sambrook et al, 1989, Molecular Cloning A Laboratory

Manual, 2nd Ed., Cold Spring Harbor Laboratory Press at 18.75), indicating the amount of target mRNA present in the original sample.

In another embodiment of this method, several different targets can be simultaneously analyzed. Two or more target-specific primers can be used in the first-strand synthesis reaction. Identifiable and distinct peptide-labeled primers complementary to the target-specific primers can be used in the PCR reaction. In this embodiment, the primers involved are chosen to be compatible in terms of their melting temperatures (Tm's) and propensities for secondary structure formation.

5.6 CHOOSING INPUT PHENOTYPES

The input phenotypes represented by cDNA libraries employed in the methods of this invention can be chosen as desired by one skilled in the art. In addition, one can use methods disclosed in co-pending United States Patent Application entitled, "Method For Identifying Genes Underlying Defined Phenotypes" by Iris, F. and Pourny, J-L., Serial No. 09/007,905, filing date January 15, 1998 for choosing in phenotypes. This co-pending application is incorporated herein by reference in its entirety.

5.7 METHODS AND PRODUCTS OF USE WITH THE INVENTION

5.7.1 DNA AMPLIFICATION

The polymerase chain reaction (PCR) is used in connection with the invention to amplify a desired sequence from a source (*e.g.*, a tissue sample, a genomic or cDNA library). Oligonucleotide primers representing known sequences can be used as primers in PCR. PCR is typically carried out by use of a thermal cycler (*e.g.*, from Perkin-Elmer Cetus) and a thermostable polymerase (*e.g.*, Gene Amp™ brand of Taq polymerase). The nucleic acid template to be amplified may include but is not limited to mRNA, cDNA or genomic DNA from any species. The PCR amplification method is well known in the art (*see, e.g.*, U.S. Patent Nos. 4,683,202, 4,683,195 and 4,889,818; Gyllenstein et al., 1988, Proc. Nat'l. Acad. Sci. U.S.A. 85, 7652-7656; Ochman et al., 1988, Genetics 120, 621-623; Loh et al., 1989, Science 243, 217-220).

Any prokaryotic cell, eukaryotic cell, or virus, can serve as the nucleic acid source. For example, nucleic acid sequences may be obtained from the following sources:

human, porcine, bovine, feline, avian, equine, canine, insect (*e.g., Drosophila*), invertebrate (*e.g., C. elegans*), plant, *etc.* The DNA may be obtained by standard procedures known in the art (*see, e.g.*, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II).

5.7.2 ADJUSTING STRINGENCY

Other methods available for use in connection with the methods of this invention include nucleic acid hybridization under low, moderate, or high stringency conditions (*e.g.*, Northern and Southern blotting). Methods for adjustment of hybridization stringency are well known in the art (*see, e.g.*, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; *see, also*, Ausubel et al., eds., in the Current Protocols in Molecular Biology series of laboratory technique manuals, 1987-1994 Current Protocols, 1994-1997 John Wiley and Sons, Inc.; *see, especially*, Dyson, N.J., 1991, Immobilization of nucleic acids and hybridization analysis, In: Essential Molecular Biology: A Practical Approach, Vol. 2, T.A. Brown, ed., pp. 111-156, IRL Press at Oxford University Press, Oxford, U.K.; each of which is incorporated by reference herein in its entirety). Salt concentration, melting temperature, the absence or presence of denaturants, and the type and length of nucleic acid to be hybridized (*e.g.*, DNA, RNA, PNA) are some of the variables considered when adjusting the stringency of a particular hybridization reaction according to methods known in the art.

Conditions of low stringency, by way of example and not limitation, may be as follows (*see, also*, Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. U.S.A. 78, 6789-6792). Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl

(pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and re-exposed to film.

Conditions of high stringency, by way of example and not limitation, may be as follows. Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 min before autoradiography.

5.7.3 OLIGONUCLEOTIDE ANALOGS

Nucleic acids used in conjunction with the device of the invention are often oligonucleotides ranging from 10 to about 50 nucleotides in length. In specific aspects, an oligonucleotide is 10 nucleotides, 15 nucleotides, 20 nucleotides or 50 nucleotides in length. An oligonucleotide can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, or single-stranded or double-stranded, or partially double-stranded. An oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, or a combination thereof. An oligonucleotide may include other appending groups, such as biotin, fluorophores, or peptides.

An oligonucleotide may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), pseudouracil, queosine, 2-thiocytosine,

5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, and 2,6-diaminopurine.

An oligonucleotide may comprise at least one modified phosphate backbone selected from the group including but not limited to a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

An oligonucleotide or derivative thereof used in conjunction with the methods of this invention may be synthesized using any method known in the art, *e.g.*, by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, *etc.*). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16, 3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Nat'l Acad. Sci. U.S.A. 85, 7448-7451), *etc.* An oligonucleotide may be an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (see Gautier et al., 1987, Nucl. Acids Res. 15, 6625-6641).

Oligonucleotides may be synthesized using any method known in the art (*e.g.*, standard phosphoramidite chemistry on an Applied Biosystems 392/394 DNA synthesizer). Further, reagents for synthesis may be obtained from any one of many commercial suppliers.

Spacer phosphoramidite molecules may be used during oligonucleotide synthesis, *e.g.*, to bridge sections of oligonucleotides where base pairing is undesired or to position labels or tags away from an oligonucleotide portion undergoing base pairing. The spacer length can be varied by consecutive additions of spacer phosphoramidites. Spacer phosphoramidite molecules may be used as 5'- or 3'- oligonucleotide modifiers. Such spacers include Spacer Phosphoramidite 9 (*i.e.*, 9-O-Dimethoxytrityl-triethyleneglycol, 1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite, and Spacer Phosphoramidite 18 (*i.e.*, 18-O-Dimethoxytrityl-hexaethyleneglycol, 1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite), both available from Glen Research (Sterling, Virginia).

Other spacers are available for use in standard oligonucleotide synthesis. For example, Spacer Phosphoramidite C3 and dSpacer Phosphoramidite can be used to destabilize undesirable self-hybridization events within capture oligonucleotides or to destabilize false hybridization events between incorrectly-matched template/probe complexes. Such spacers, when positioned at the 3' end of an oligonucleotide, will also prevent incorrect extension products from being generated when included in a PCR reaction mixture.

One spacer available from Glen Research, Spacer Phosphoramidite C3 (*i.e.*, 3-O-Dimethoxytrityl-propyl-1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite), can be added to substitute for an unknown base within an oligonucleotide sequence.

A branching spacer may be used as one method to increase label incorporation into an oligonucleotide. Such a branching spacer may also be used to increase a detectable signal by hybridization through multiply branched capture probes or PCR primers. Branching spacers are available commercially, *e.g.*, from Glen Research.

Biotinylated oligonucleotides are well known in the art. An oligonucleotide may be biotinylated using a biotin-NHS ester procedure. Alternatively, biotin may be attached during oligonucleotide synthesis using a biotin phosphoramidite (Cocuzza, 1989, Tetrahed. Lett. 30, 6287-6290). One such biotin phosphoramidite available from Glen Research is 1-Dimethoxytrityloxy-2-(N-biotinyl-4-aminobutyl)-propyl-3-O-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite. This compound also has a branch point to allow further additions. The branched spacer used in this biotin phosphoramidite has been described by Nelson et al. (1992, Nucl. Acids Res. 20, 6253-6259).

Another 5'-biotin phosphoramidite, namely [1-N-(4,4'-Dimethoxytrityl)-biotinyl-6-aminohexyl]-2-cyanoethyl-(N,N-diisopropyl)-phosphoramidite, may be used to biotinylate an oligonucleotide. This compound is sold by Glen Research under license from Zeneca PLC.

Fluorescent dyes may also be incorporated into an oligonucleotide using dye-labeled phosphoramidites. Two such labels are 5'-Hexachloro-Fluorescein Phosphoramidite (HEX), and 5'-Tetrachloro-Fluorescein Phosphoramidite (TET), both available from Glen Research.

5.7.4 PRODUCTION OF LABELED OLIGONUCLEOTIDES

Oligonucleotides may be labeled with a wide variety of lables for use in the various embodiments of the invention. For example, European Patent Publication No. EP 0370 694 A2, entitled, "Diagnostic Kit and Method Using a Solid Phase Capture Means For Detecting Nucleic Acid", by Burdick and Oakes, publication date May 30, 1990, discloses methods of linking labels to oligonucleotides.

Methods of attaching peptides to oligonucleotides are well known to those with ordinary skill in the art, *e.g.*, *see*, 1) Soukchareun S. et al., Preparation and characterization of antisense oligonucleotide-peptide hybrids containing viral fusion peptides. Bioconjug. Chem.,1995, 6(1):43-53; 2) Tung CH, et al., Preparation of oligonucleotide-peptide conjugates. Bioconjug. Chem.,1991, 2(6):464-465; 3) Bruick RK, et al., Template-directed ligation of peptides to oligonucleotides. Chem. Biol.,1996, 3(1):49-56; 4) Tung CH, et al., Dual-specificity interaction of HIV-1 TAR RNA with Tat peptide-oligonucleotide conjugates. Bioconjug. Chem.,1995, 6(3):292-295; 5) Robles J., et al., Synthesis and Enzymatic Stability of Phosphodiester-Linked Peptide-Oligonucleotide Hybrids, Bioconjug. Chem., 1997, 8(6):785-788 ; and 6) Rajur S.B., et al., Covalent Protein-Oligonucleotide Conjugates for Efficient Delivery of Antisense Molecules, Bioconjug. Chem.,1997, 8(6):935-940.

Oligonucleotides linked to various peptides for use in the methods of this invention may be obtained for example, from Cybergene S.A. (11 rue Claude Bernard, z1 nord, 35400, Saint Mallo, France) and Glen Research (22825 Davis Drive, Sterling, Virginia 20164). Further information from Glen Research can be obtained through their web site (www.glenres.com).

One specific method for linking a peptide to an oligonucleotide recommended by Glen Research is as follows (see also, www.glenres.com). A heterobifunctional crosslinking reagent is used to link a synthetic peptide having an N-terminal lysine residue to a 5'-thiol-modified oligonucleotide. Such a crosslinking reagent is N-maleimido-6-aminocaproyl-(2'-nitro, 4'-sulfonic acid) phenyl ester (mal-sac-HNSA). The sodium salt of mal-sac-HNSA is available from Bachem Bioscience. Conveniently, reaction of the mal-sac-HNSA crosslinker with an amino group releases a dianion phenolate (i.e. 1-hydroxy-2-nitro-4-benzene sulfonic acid). This dianion phenolate is also a yellow

chromophore. The chromophore feature provides (i) a means for quantifying the extent of completion of the coupling reaction (where greater yellow color intensity corresponds to a more complete coupling reaction), and (ii) an aid in monitoring the extent of separation of an activated peptide (i.e. a peptide crosslinked to mal-sac-HNSA and ready for contacting with a 5'-thiol-modified oligonucleotide) from free crosslinking reagent during gel filtration.

The specific steps employed when using a mal-sac-HNSA crosslinker may be as follows. First, a peptide is synthesized having an N-terminal lysine. Alternatively, a peptide having an internal lysine may be used since the lysine epsilon amino group is actually more reactive than the lysine alpha amino group. Second, an oligonucleotide is synthesized having a 5'-thiol group using methods known in the art. Third, the peptide is reacted with an excess of mal-sac-HNSA in a sodium phosphate buffer (pH 7.1). Fourth, the peptide-mal-sac conjugate is separated from free crosslinker and the buffer is exchanged to sodium phosphate (pH 6) using a gel filtration column (e.g. NAP-5, Pharmacia, Uppsala, Sweden). Fifth, a thiol-modified oligonucleotide is activated, desalted and buffer-exchanged to sodium phosphate (pH 6) on a gel filtration column. Sixth, the activated peptide is reacted with the thiol-modified oligonucleotide. Finally, the peptide-oligonucleotide conjugate is purified by ion exchange chromatography (e.g. Nucleogen DEAE-500-10 or equivalent). The elution order from the ion exchange column is as follows: free peptide first, peptide-labeled oligonucleotide next, and free oligonucleotide last.

5.7.5 ANTIBODIES AND PEPTIDES

Antibodies of use with the methods of this invention include any antibodies known in the art. Such antibodies may be used, for example, to manipulate the nucleic acids of interest. In this regard, a nucleic acid may be manipluated by antibody binding to the nucleic acid itself or to an antigen (*e.g.*, a protein, peptide or hapten) which is bound (either covalently or non-covalently) to the nucleic acid. In a preferred embodiment, nucleic acids are manipulated using peptide antigens covalently attached to PCR primers. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric and humanized antibodies, as described below. Further, single chain antibodies, Fab fragments

and $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above may also be used.

Polyclonal antibodies which may be used with the invention are heterogeneous populations of antibody molecules derived from the sera of immunized animals. Various procedures well known in the art may be used for the production of polyclonal antibodies to an antigen-of-interest. For example, the production of polyclonal antibodies, various host animals can be immunized by injection with an antigen of interest or derivative thereof, including but not limited to rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum. Such adjuvants are also well known in the art.

Monoclonal antibodies which may be used with the invention are homogeneous populations of antibodies to a particular antigen. A monoclonal antibody (mAb) to an antigen-of-interest can be prepared by using any technique known in the art which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Kohler and Milstein (1975, Nature 256, 495-497), and the more recent human B cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4, 72), and the EBV-hybridoma technique (Cole et al., 1985, Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAbs of use in this invention may be cultivated *in vitro* or *in vivo*.

Monoclonal antibodies which may be used with the methods of the invention include but are not limited to human monoclonal antibodies. Human monoclonal antibodies may be made by any of numerous techniques known in the art (*e.g.*, Teng et al., 1983, Proc. Nat'l Acad. Sci. U.S.A. 80, 7308-7312; Kozbor et al., 1983, Immunology Today 4, 72-79; Olsson et al., 1982, Meth. Enzymol. 92, 3-16).

A chimeric antibody may be used with the methods of the invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. Various techniques are available for the production of such chimeric antibodies (*see, e.g.*, Morrison et al., 1984, Proc. Nat'l Acad. Sci. U.S.A. 81, 6851-6855; Neuberger et al., 1984, Nature, 312, 604-608; Takeda et al., 1985, Nature, 314, 452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity.

A humanized monoclonal antibody may be used with the methods of the invention. Briefly, humanized antibodies are antibody molecules from non-human species having one or more complementarily determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. Various techniques have been developed for the production of humanized antibodies (*see, e.g.*, Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety). An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, referred to as complementarily determining regions (CDRs). The extent of the framework region and CDRs have been precisely defined (*see*, Kabat et al., 1983, Sequences of proteins of immunological interest, U.S. Department of Health and Human Services).

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778; Bird, 1988, Science 242, 423-426; Huston et al., 1988, Proc. Nat'l. Acad. Sci. U.S.A. 85, 5879-5883; and Ward et al., 1989, Nature 334, 544-546) can be adapted to produce single chain antibodies useful in the device of the invention. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region together via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, Science, 246,

1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Further general methods of antibody production and use are suitable for use in connection with the methods of the invention. For example see Harlow and Lane, 1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, which is incorporated herein by reference in its entirety.

The single-lettter amino acid code corresponds to the three-letter amino acid code of the Sequence Listing set forth hereinbelow, as follows: A, Ala; R, Arg; N, Asn; D, Asp; B, Asx; C, Cys; Q, Gln; E, Glu; Z, Glx; G, Gly; H, His; I, Ile; L, Leu; K, Lys; M, Met; F, Phe; P, Pro; S, Ser; T, Thr; W, Trp; Y, Tyr; and V, Val.

Suitable antibodies for use with the methods of this invention include the following, available from Affinity Bioreagents, Inc., 79, rue des Morillons, 75015, Paris, France.

1) Catalog No. PA 1-047 (affinity-purified rabbit IgG). The corresponding peptide recognized by this Ab is KFSREKKAAKT (SEQ ID NO:1).

2) Catalog No. PA 1-039 (affinity-purified rabbit immunogobins). The corresponding peptide recognized by this Ab is DQKRYHEDIFG (SEQ ID NO:2).

3) Catalog No. PA 1-036 (purified rabbit IgG). The corresponding peptide recognized by the Ab is DLKEEKDINNNVKKT (SEQ ID NO:3).

4) Catalog No. PA 1-014 (purified rabbit antibody). The corresponding peptide recognized by this Ab is CTGEEDTSE (SEQ ID NO: 4).

5) Catalog No. PA 3-013 (affinity purified IgG). The corresponding peptide recognized by this Ab is PEETQTQDQPM (SEQ ID NO:5).

6) Catalog No. PA 1-815 (rabbit anti-serum). The corresponding peptide recognized by this Ab is QKSDQGVEGPGAT (SEQ ID NO:6).

7) Catalog No. PA 3-034 (rabbit polyclonal serum IgG). The corresponding peptide recognized by this Ab is DIGQSIKKFSKV (SEQ ID NO:7). This polyclonal antibody will also recognize QRADSLSSHL (SEQ ID No:8).

In addition, antibodies for use with the methods of this invention may be obtained from Medical & Biological Laboratories Co., Ltd., 440 Arsenal Street, Watertown, Massachusetts 02171, U.S.A.

These include the following:

1) Code No. 561 (Rabbit IgG from anti-serum). The corresponding peptide recognized is YPYDVPDYA (SEQ ID NO:9).

2) Code No. 562 (Rabbit IgG from anti-serum). The corresponding peptide recognized is EQKLISEEDL (SEQ ID NO:10).

3) Code No. 563 (Rabbit IgG from anti-serum). The corresponding peptide recognized is YTDIEMNKLGK (SEQ ID NO:11).

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed since these embodiments are intended as illustration of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Throughout this application various references are cited, the contents of each of which is hereby incorporated by reference into the present application in its entirety.

WE CLAIM:

1. A method of sorting a mixture of nucleic acids derived from a plurality of cDNA libraries comprising:

(a) labeling DNA from each of the plurality of cDNA libraries by polymerase chain reaction using oligonucleotide primers having a label distinguishable to each library;

(b) contacting DNA labeled in step (a) with a first said label with DNA labeled in step (a) with a different said label under conditions such that hybridization can occur; and

(c) sorting DNA contacted in step (b) using one or more molecules, each molecule being capable of binding one of the labels distinguishable to each library.

2. The method of claim 1 wherein the label distinguishable to each library is a 5'-peptide label.

3. The method of claim 1 wherein the label distinguishable to one library is biotin.

4. The method of claim 1 wherein at least one of the one or more molecules is an antibody.

5. The method of claim 1 wherein the oligonucleotide primers prime polymerase chain reaction from vector sequences common to the plurality of cDNA libraries.

6. The method of claim 1 wherein said sorting comprises:

(d) denaturing hybrid DNA strands resulting from step (b);

(e) contacting single strands denatured in (d) with single strand binding protein to prevent strand reannealing; and

(f) contacting single strand binding protein coated single strands formed in (e) with one or more molecules each molecule being capable of binding one of the labels distinguishable to each library.

7. The method of claim 1 or claim 6 wherein at least one of the one or more molecules is an antibody.

8. A method of cDNA library comparison comprising:

(a) labeling DNA from a first cDNA population by polymerase chain reaction using oligonucleotide primers having a first 5'-peptide label;

(b) labeling DNA from a second cDNA population by polymerase chain reaction using oligonucleotide primers having a second 5'-peptide label;

(c) contacting DNA labeled in step (a) with DNA labeled in step (b) under conditions such that hybridization can occur; and

(d) separating DNA having the first and second 5' peptide labels from DNA having only the first or the second 5' peptide label.

9. The method of claim 8 wherein the first cDNA population is from one or more cells or an organism subjected to a first condition and the second cDNA population is from one or more cells or an organism of the same type not subjected to said first condition.

10. The method of claim 8 wherein the first cDNA population is from one or more cells or an organism subjected to a first condition and the second cDNA population is from one or more cells or an organism of the same type subjected to a second condition.

11. The method of claim 8 wherein the first and second cDNA populations are from cells or organisms that differ phenotypically.

12. The method of claim 8 wherein the nucleotide sequences of the oligonucleotide primer pair having the first 5'-peptide label and the nucleotide sequences of the oligonucleotide primer pair having the second 5'-peptide label are the same.

13. A method of monitoring gene expression comprising:

(a) contacting mRNA from a cell with an RNA-dependent DNA polymerase and a 5'-dephosphorylated target-specific primer;

(b) contacting any DNA:RNA hybrid molecules synthesized in step (a) with a nuclease to remove single-stranded RNA extensions;

(c) after step (b) ligating the DNA:RNA hybrids molecules to a partly double-stranded phosphorylated primer;

(d) labeling products ligated in step (c) by polymerase chain reaction with a first primer complementary to the target-specific primer used in step (a), said first primer being labeled with a first label, and a second primer complementary to one strand of the double-stranded phosphorylated primer used in (c), said second primer being labeled with a second label that is distinguishable from said first label;

(e) contacting the polymerase chain reaction products labeled in step (d) with one or more molecules immobilized on a solid support capable of binding the first label;

(f) washing the solid support; and

(g) contacting the support washed in step (f) with one or more molecules capable of binding the second label.

14. The method of claim 13 wherein the nuclease is mung-bean nuclease.

15. The method of claim 13 wherein the partly double-stranded phosphorylated primer is an M13 forward sequencing primer.

16. The method of claim 13 wherein the first label is a peptide label.

17. The method of claim 13 wherein the second label is biotin.

18. The method of claim 13 wherein at least one of the one or more molecules in step (e) is an antibody.

19. The method of claim 13 wherein at least one of the one or more molecules in step (g) is streptavidin-linked horseradish peroxidase.

20. A method of identification of cDNA inserts represented in a first cDNA library and not represented in a plurality of other cDNA libraries comprising:

(a) labeling DNA inserts from each cDNA library by polymerase chain reaction using oligonucleotide primers having a label unique to each library;

(b) hybridizing DNA labeled in step (a);

(c) contacting DNA hybridized in step (b) with a plurality of immobilized antibodies capable of recognizing the label unique to each of the plurality of other cDNA libraries but not the label unique to the first cDNA library; and

(d) recovering DNA which is not bound by the plurality of immobilized antibodies.

21. The method of claim 20, wherein DNA hybridized from each of the plurality of other cDNA libraries is in excess relative to the first cDNA library.

22. The method of claim 21, wherein the excess is from a 2-fold to a 100-fold excess.

23. The method of claim 21, wherein the excess is from a 2.5-fold to a 10-fold excess.

24. The method of claim 21, wherein the excess is a 3-fold excess.

25. The method of claim 20, wherein the label unique to each library is a peptide label.

26. The method of claim 25, wherein the peptide label is from 3 to 12 amino acid residues.

27. The method of claim 20, wherein the label unique to each library is a thermophilic protein label.

28. The method of claim 20, wherein each of the plurality of antibodies in step (c) is immobilized on a separate affinity column.

29. The method of claim 28, wherein the separate affinity columns are physically linked in series in any order.

30. The method of claim 29, wherein column flow-through is applied to the separate, physically-linked affinity columns one or more times.

31. The method of claim 29, wherein column flow-through is applied to the separate, physically-linked affinity columns three times.

32. The method of claim 20, wherein DNA recovered in step (d) is further contacted with an antibody specific for the label unique to the first cDNA library.

33. The method of claim 32, wherein DNA retained by the antibody specific for the label unique to the first cDNA library is recovered and cloned.

34. A method of identification of cDNA inserts represented in a first cDNA library and in a second cDNA library, and not represented in a plurality of other cDNA libraries, comprising:

(a) labeling DNA from each cDNA library by polymerase chain reaction using oligonucleotide primers having a label unique to each library;

(b) hybridizing DNA labeled in step (a);

(c) contacting DNA hybridized in step (b) with a plurality of immobilized antibodies capable of recognizing the label unique to each of the plurality of other cDNA libraries but not the label unique to the first cDNA library or the second cDNA library; and

(d) recovering DNA which is not bound by the plurality of immobilized antibodies.

35. The method of claim 34, wherein DNA hybridized from each of the plurality of other cDNA libraries is in excess relative to the first and second cDNA libraries.

36. The method of claim 35, wherein the excess is from a 2-fold to a 100-fold excess.

37. The method of claim 35, wherein the excess is from a 2.5-fold to a 10-fold excess.

38. The method of claim 35, wherein the excess is a 3-fold excess.

39. The method of claim 34, wherein the label unique to each library is a peptide label.

40. The method of claim 39, wherein the peptide label is from 3 to 12 amino acid residues.

41. The method of claim 34, wherein the label unique to each library is a thermophilic protein label.

42. The method of claim 34, wherein each of the plurality of antibodies in step (c) is immobilized on a separate affinity column.

43. The method of claim 42, wherein the separate affinity columns are physically linked in series in any order.

44. The method of claim 43, wherein column flow-through is applied to the separate, physically-linked affinity columns one or more times.

45. The method of claim 43, wherein flow-through is applied to the separate, physically-linked affinity columns three times.

46. The method of claim 34, wherein DNA recovered in step (d) is further contacted with an antibody specific for the label unique to the first cDNA library or the label unique to the second cDNA library so as to concentrate cDNA fragments specific to the first cDNA library and the second cDNA library.

47. The method of claim 46, wherein the concentrated cDNA fragments specific to the first cDNA library and the second cDNA library are recovered and cloned.

48. The method of claim 47, wherein the concentrated cDNA fragments specific to the first cDNA library and the second cDNA library are separated.

49. The method of claim 48, wherein separation is carried out by denaturation, coating with single strand binding protein, and contacting with an antibody specific for the label unique to the first cDNA library or the second cDNA library.

50. A method for matrix analysis of a plurality of cDNA libraries comprising:

(a) labeling cDNA inserts from each of the plurality with a distinguishable label;

(b) hybridizing cDNA inserts labeled in step (a);

(c) contacting cDNA inserts hybridized in step (b) with an affinity column capable of binding a distinguishable label; and

(d) eluting the affinity column.

51. The method of Claim 50, wherein the distinguishable label is a peptide label, and the step of labelling comprises priming polymerase chain reaction from cDNA library vector sequences by use of an oligonucleotide primer pair having said peptide label attached to the 5 ends of said primer pair.

52. The method of claim 50, wherein the labeled cDNA fragments from each library are hybridized in equal proportions.

53. The method of claim 50, wherein the affinity column capable of binding a distinguishable label is an antibody affinity column.

54. The method of claim 53, wherein the antibody affinity column is eluted with a pH gradient.

55. The method of claim 50, wherein eluted DNA is denatured to separate strands originating from two different libraries.

56. The method of claim 55, wherein denatured strands are isolated by: (a) coating with single-strand binding protein; and (b) contacting with an affinity column capable of binding a distinguishable label.

PHASE I
PHASE II
PHASE III
GROUP "A"
GROUP "B"
GROUP "C"
GROUP "D"
GROUP "E"
A:B
A:C
A:D
A:E
B
C
D
E
B:B + B–
C:C + C–
D:D + D–
E:E + E–
B:A
B:C
B:D
B:E
A
C
D
E
A:A + A–
B:B + B–
D:D + D–
E:E + E–
C:A
C:B
C:D
C:E
A
B
C
D
E
A:A + A–
B:B + B–
D:D + D–
E:E + E–
D:A
D:B
D:C
D:E
A
B
C
E
A:A + A–
B:B + B–
C:C + C–
E:E + E–
E:A
E:B
E:C
E:D
A
B
C
D
A:A + A–
B:B + B–
C:C + C–
D:D + D–
TRANSCRIPTS SHARED BETWEEN LIBRARIES
ISOLATION OF SHARED TRANSCRIPTS
REMOVAL OF HOMOLOGOUS SHARED TRANSCRIPTS

FIG. 1

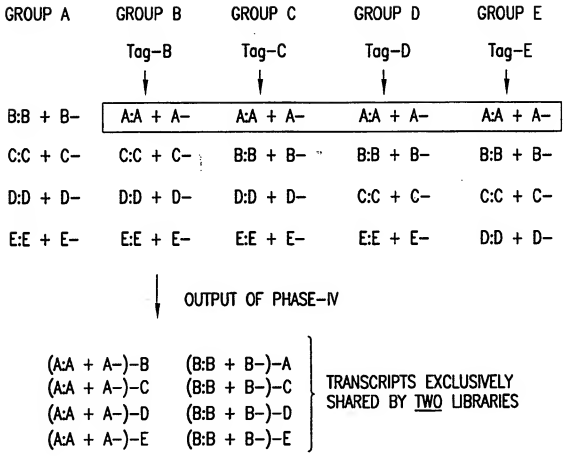
GROUP A
GROUP B
GROUP C
GROUP D
GROUP E
Tag–B
Tag–C
Tag–D
Tag–E
B:B + B–
A:A + A–
A:A + A–
A:A + A–
A:A + A–
C:C + C–
C:C + C–
B:B + B–
B:B + B–
B:B + B–
D:D + D–
D:D + D–
D:D + D–
C:C + C–
C:C + C–
E:E + E–
E:E + E–
E:E + E–
E:E + E–
D:D + D–
OUTPUT OF PHASE–IV
(A:A + A–)–B
(A:A + A–)–C
(A:A + A–)–D
(A:A + A–)–E
(B:B + B–)–A
(B:B + B–)–C
(B:B + B–)–D
(B:B + B–)–E
TRANSCRIPTS EXCLUSIVELY SHARED BY TWO LIBRARIES

FIG. 2

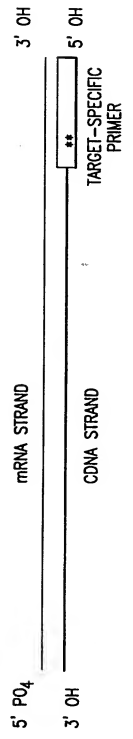
3' OH
5' OH
**
TARGET-SPECIFIC PRIMER
mRNA STRAND
CDNA STRAND
5' PO4
3' OH

FIG. 3

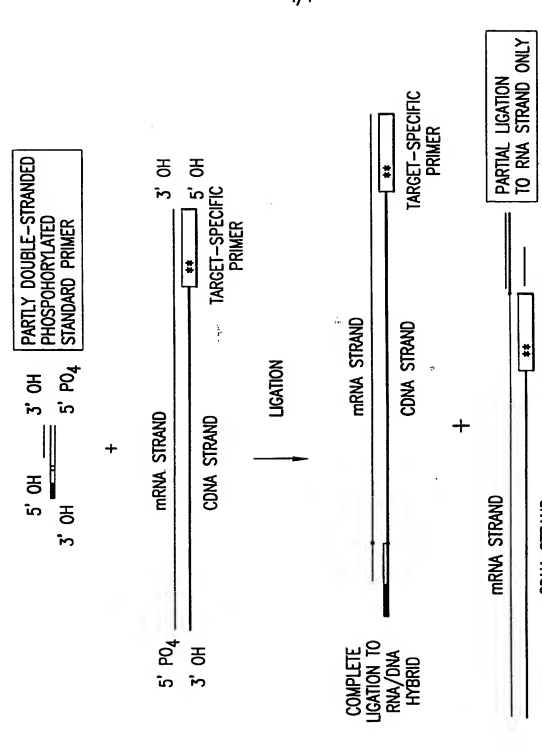
5' OH
3' OH
3' OH
5' PO4
PARTLY DOUBLE-STRANDED PHOSPHORYLATED STANDARD PRIMER
+
5' PO4
3' OH
mRNA STRAND
CDNA STRAND
3' OH
5' OH
**
TARGET-SPECIFIC PRIMER
LIGATION
COMPLETE LIGATION TO RNA/DNA HYBRID
mRNA STRAND
CDNA STRAND
**
TARGET-SPECIFIC PRIMER
+
mRNA STRAND
CDNA STRAND
**
TARGET-SPECIFIC PRIMER
PARTIAL LIGATION TO RNA STRAND ONLY

FIG. 4

SEQUENCE LISTING

<110> VALIGENE CORPORATION

<120> METHODS FOR MANIPULATING COMPLEX NUCLEIC ACID POPULATIONS USING PEPTIDE-LABELED OLIGONUCLEOTIDES

<130> 9408-025-228

<140> PCT/US99/23906
<141> 1999-10-15

<160> 11

<170> PatentIn Ver. 2.0

<210> 1
<211> 11
<212> PRT
<213> Oryctolagus cuniculus

<400> 1
Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys Thr
1 5 10

<210> 2
<211> 11
<212> PRT
<213> Oryctolagus cuniculus

<400> 2
Asp Gln Lys Arg Tyr His Glu Asp Ile Phe Gly
1 5 10

<210> 3
<211> 15
<212> PRT
<213> Oryctolagus cuniculus

<400> 3
Asp Leu Lys Glu Glu Lys Asp Ile Asn Asn Asn Val Lys Lys Thr
1 5 10 15

<210> 4
<211> 9
<212> PRT
<213> Oryctolagus cuniculus

<400> 4
Cys Thr Gly Glu Glu Asp Thr Ser Glu
1 5

<210> 5
<211> 11
<212> PRT
<213> Oryctolagus cuniculus

<400> 5
Pro Glu Glu Thr Gln Thr Gln Asp Gln Pro Met
1 5 10

<210> 6
<211> 13
<212> PRT
<213> Oryctolagus cuniculus

<400> 6
Gln Lys Ser Asp Gln Gly Val Glu Gly Pro Gly Ala Thr
1 5 10

<210> 7
<211> 12
<212> PRT
<213> Oryctolagus cuniculus

<400> 7
Asp Ile Gly Gln Ser Ile Lys Lys Phe Ser Lys Val
1 5 10

<210> 8
<211> 10
<212> PRT
<213> Oryctolagus cuniculus

<400> 8
Gln Arg Ala Asp Ser Leu Ser Ser His Leu
1 5 10

<210> 9
<211> 9
<212> PRT
<213> Oryctolagus cuniculus

<400> 9
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 10
<211> 10
<212> PRT
<213> Oryctolagus cuniculus

<400> 10
Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 11
<211> 11
<212> PRT
<213> Oryctolagus cuniculus

<400> 11
Tyr Thr Asp Ile Glu Met Asn Lys Leu Gly Lys
1 5 10

INTERNATIONAL SEARCH REPORT

Inte ional Application No
PCT/US 99/23906

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DUGUID J R ET AL: "LIBRARY SUBTRACTION OF IN VITRO CDNA LIBRARIES TO IDENTIFY DIFFERENTIALLY EXPRESSED GENES IN SCRAPIE INFECTION" NUCLEIC ACIDS RESEARCH,GB,OXFORD UNIVERSITY PRESS, SURREY, vol. 18, no. 9, 11 May 1990 (1990-05-11), pages 2789-2792, XP000371880 ISSN: 0305-1048 see whole doc. esp. fig. 1	1
X	WO 89 01526 A (GENELABS INC) 23 February 1989 (1989-02-23) see whole doc. esp. claims (8,9)	1,8,20, 34,50

-/---

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 March 2000	23/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Müller, F

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte ional Application No
PCT/US 99/23906

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 43443 A (BOEHRINGER MANNHEIM GMBH ;SCHIBLER UELI (CH); LAVERY DANIEL (CH)) 20 November 1997 (1997-11-20) the whole document	1,8,20, 34,50
A	US 5 804 382 A (YANG MEIHENG ET AL) 8 September 1998 (1998-09-08) the whole document	
P,X	WO 99 36575 A (VALIGENE CORP) 22 July 1999 (1999-07-22) see whole doc, esp. claim 38 and p.18, l.21 -p.19, l.7	1,8,34, 50

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Serial Application No
PCT/US 99/23906

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 8901526	A	23-02-1989	AU 2318288 A		09-03-1989
WO 9743443	A	20-11-1997	EP 0839214 A JP 11509427 T		06-05-1998 24-08-1999
US 5804382	A	08-09-1998	NONE		
WO 9936575	A	22-07-1999	AU 2234699 A		02-08-1999